Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.28, NO.5, Oct. 2018 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2018.28.5.1169

사전유입 에이전트가 발생하는 이상트래픽 탐지 방안

조 영 민,† 권 헌 영‡
고려대학교 정보보호대학원

Detection of Abnormal Traffic by Pre-Inflow Agent

Young Min Cho,† Hun Yeong Kwon‡
Graduate School of Information Security, Korea University

요 약

현대 사회는 급격한 디지털 트랜스포메이션 시대라 할 수 있다. 이러한 디지털 중심의 비즈니스 확산은 기업과 개인에게 편리함과 효율성을 제공하지만 그만큼 사이버 위협은 증가하고 있다. 특히 사이버 공격은 점차 지능화, 정밀화되면서 다양화 되고 있으며 이러한 공격이 발각되지 않도록 다양한 방법을 시도하고 있다. 따라서 이러한 공격에 대응 하는 것이 점점 어려워지고 있는 현실이다. 사이버킬체인(Cyber Kill Chain) 개념에 따르면 공격자는 여러 단계에 걸쳐 목적을 달성하기 위해 침투하게 되는데 우리는 이러한 여러 단계중 하나를 탐지하여 공격을 무력화하는 것이 목적이다. 본 논문에서는 시스템의 오류 또는 사용자의 실수 등 다양한 원인으로 사전에 악성행위를 실행하는 에이전트가(agent) 유입되었다고 가정하고, 이러한 에이전트가 외부의 공격자와 접속하기 위해 발생시키는 이상트래픽을 탐지하는 방안을 제안하고자 한다.

ABSTRACT

Modern society is a period of rapid digital transformation. This digital-centric business proliferation offers convenience and efficiency to businesses and individuals, but cyber threats are increasing. In particular, cyber attacks are becoming more and more intelligent and precise, and various attempts have been made to prevent these attacks from being discovered. Therefore, it is increasingly difficult to respond to such attacks. According to the cyber kill chain concept, the attacker penetrates to achieve the goal in several stages. We aim to detect one of these stages and neutralize the attack. In this paper, we propose a method to detect anomalous traffic caused by an agent attacking an external attacker, assuming that an agent executing a malicious action has been introduced in advance due to various reasons such as a system error or a user's mistake.

Keywords: abnormal traffic, agent, pre-inflow, detection

I. 서 론

현대 사회는 기업이 새로운 비즈니스 모델, 제품, 서비스를 만들기 위해 디지털 역량을 기반으로 혁신적인 변화에 적응하는 디지털트랜스포메이션(Digital Transformation) 시대라 할 수 있다[1]. 이러한 디지털중심의 시대는 인간과 사회에 많은 편리함을 제공하지만 한편으로는 그에 따른 다양한 사이버 위협이 존재하고 있는 상황이다. 실제로 IDG에서는 2021년까지 사이버범죄 피해비용이 매년 6조 달러에 달할 것으로 전망했다[2]. 한편 체크포인트의 개리 킨슬리(Gary Kinsley)는 IT발전에 따른 사이버 위협을 5개의 세대로 정의했다. 1세대는 바이러스에 의한 PC공격(1980년대), 2세대는

Received(07. 17. 2018), Modified(09. 03. 2018), Accepted(09. 13. 2018)
† 주저자, ymin_cho@naver.com
‡ 교신저자, cabkwon@gmail.com(Corresponding author)

네트워크 공격으로 인한 방화벽 대응(1990년대), 3세대는 애플리케이션 공격(2000년대), 4세대 페이로드공격(2010년)에 이어, 현재는 대규모성과 복합적인 공격요소로 설명되는 5세대에 도래했다고 소개한다[3]. 이와 같이 최근의 사이버 공격은 다양한 방법을 동원하여 목적을 달성하기 위해 지속적으로 치밀해지고 복합적인 방법을 시도하고 있다.

사이버킬체인(Cyber Kill Chain)에서 언급되는 바와 같이 외부의 공격자는 위해 목적 달성을 위해 여러 단계를 거쳐 목표물에 접근하는 점진적 접근을 시도하게 된다[4]. 이때 각 단계별로 공격자는 공격 행위가 발각되지 않도록 다양한 기법을 통해 공격을 시도할 것이며, 외부(원격지)에 있는 공격자는 목표 지점의 단계별 행동을 위해 악성코드라 불리는 에이전트를 작동하여 원하는 목적을 달성하게 된다. 결국 공격자는 단계별 미션을 수행하기 위해서는 악성코드를 어떠한 형태로든 내부에 유입시키고 이를 컨트롤하는 상황이 된다.

이러한 악성코드를 통한 공격은 지속적으로 증가하고 있으며, 2009년에 비해 약10년이 지난 2018년 현재는 약35배 이상의 악성코드가 증가한 것으로 보고되고 있으며[5], 이러한 악성코드를 탐지하기 위해 다양한 연구가 진행되고 있다.

악성코드를 탐지하는 방법은 시그니처 기반(Signature Based Technique)와 행위기반(Behavior Based Technique)로 크게 구분할 수 있으며, 이를 다시 조합하면 정적분석(Static Analysis), 동적분석(Dynamic Analysis), 복합분석(Hybrid Analysis)로 구분할 수 있다. [6][7].

그러나 이러한 탐지 노력에도 불구하고 공격자들은 그들의 에이전트를 목적지에 잠입시키기 위해 제로데이(Zero-day) 공격, 난독화, Sandbox회피, 활동시간 지연 등 다양한 탐지 우회 방법을 시도한다. 이외에도 탐지 장비를 운영하는 사용자의 실수나 탐지장비의 오작동, 미인지된 네트워크 경로 등 다양한 원인에 의해 공격자의 에이전트가 내부에 유입될 수 있다. 따라서 본 논문에서는 외부 공격자(해커)의 명령을 에이전트가 기관/기업 내부에 이미 유입되었다는 전제하에 외부 공격자와 통신하는 트래픽을 탐지하는 방법을 제안하여 공격 단계(무기화, 익스플로잇 설치, 명령/제어 등의 각 단계)의 어느 한 단계라도 흐름을 차단함으로써 공격을 무력화하는 목적으로 활용하고자 한다.

II. 관련 연구

이상트래픽을 탐지하는 방법은 공격 기법이 다양화 되면서 이에 대한 연구도 여러 분야로 수행되어 왔다.

시그니처 기반 탐지 방안은 기존에 유입된 악성코드/봇의 통신 특징을 파악하여 특정 문자열 패턴이나 세션 정보를 통하여 탐지 한다[8][9]. 시그니처 기반은 정보가 조금만 변경되어도 탐지가 어렵다는 문제가 있다. 머신러닝 기반의 탐지 방법은 트래픽 내역을 분석 데이터로 활용하여 KNN, Random Forest 등의 일반적인 머신러닝 학습을 수행하는 방식이다[11][12]. 머신러닝은 모델을 만들기 위한 학습데이터가 준비되어하는 어려움이 있다.

본 연구의 주제 영역인 행위기반 탐지 방안은 통신이 발생할 때 생기는 정보 중 byte의 크기, 주로 사용되는 port번호, 발생빈도, 이전 또는 이후 발생하는 추가 행위 정보 등을 토대로 탐지한다[9][10]. 또한 발생하는 데이터의 임계치 정보를 학습하거나 통계정보를 바탕으로 이상치를 구분한다[10]. 과거 C&C서버와 통신한 내역을 인지하고 있고, 해당 트래픽을 분석한 결과 일반적이지 않은 포트정보를 사용하며 시간단위로 포트정보가 일정하게 증가시키고, 추가로 동일한 크기의 응답 크기(byte)가 발생하는 경우 이를 하나의 행위로 등록하여 유사한 행위가 발생하는지 체크하는 방식이다. 시그니처 기반의 탐지 방법에 비해 탐지할 수 있는 범위가 넓어질 수 있으나 사전에 각 트래픽 별로 행위정보를 뽑아 정리 해 놓는 프로파일링(profiling) 작업이 필요하다. 네트워크 트래픽의 Flow 정보를 프로파일링하여 기존 정보와 유사한 것을 찾아내는 방식의 연구도 진행되었다[13]. 이 작업의 경우 행위 패턴을 벗어날 경우 탐지가 어렵다.

네트워크 정보에서 얻을 수 있는 데이터를 활용한

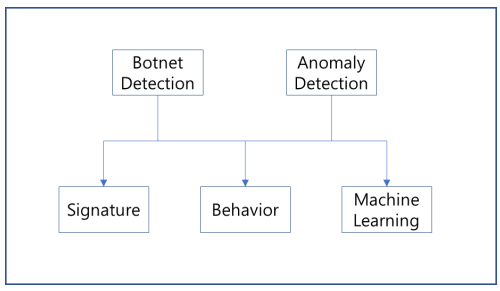

Fig. 1. Related Research Field

또 다른 방법으로 로그 데이터의 이동 평균과 3-시그마를 사용한 연구가 진행되었다[14]. 시간 데이터의 일정기간 평균을 계산하고 가중치를 부여하여 이상치를 찾아내는 방식으로, 시간 정보가 규칙성을 벗어나면 탐지가 어렵다. 기존 연구에서는 현재 또는 수분 이내와 같은 짧은 시간 내 발생하는 이벤트를 대상으로 하고 있으나, 본 논문에서는 하루 또는 일주일 등 특정 기간 전체를 대상으로 기간 내 발생한 숨겨진 위협을 보다 실용적으로 탐지하는 방안을 제시하며, 특히 시간 정보의 평균성에 의미가 없는 경우, 즉 규칙적이 아닌 접속 상황에서도 반복 접속 탐지가 가능한 방안을 제시하고자 한다.

III. 이상트래픽 탐지 방안

3.1 전제 사항

본 논문은 현재 발생하는 외부에서 사내 시스템으로 발생하는 공격 트래픽을 찾고자하는 것이 목적이 아니라, 공격자의 지령을 받을 수 있는 에이전트와 외부 공격자의 http 통신 내역 탐지를 목적으로 한다. 따라서 아래와 같은 몇 가지 사항을 전제 한다.

첫째, 악성코드 탐지 장비의 미탐, 사용자의 실수, 장비의 오작동, 알 수 없는 예외 경로 등의 다양한 원인으로 에이전트가 이미 유입되었다. 따라서 언제 어떻게 유입되었는지 알 수 없는 에이전트를 탐지하기 위해 분석 대상 기간의 전체 데이터를 대상으로 한다.

둘째, 유입된 에이전트는 외부 지령을 받아 추가적인 행위(킬체인 상의 다음 단계 진행)를 수행하려 할 것이고, 이를 위해 외부와 반복적인 접속을 시도할 것이다.

셋째, 접속 시도 중 탐지될 경우를 가정하여 하나의 접속이 아닌 여러 개의 외부 목적지로 다중 접속을 시도할 것이다.

넷째, 진행 과정 중 탐지되어 차단되더라도 공격자의 지령을 받거나 추가 악성코드를 다운로드 하기 위해 외부와 접속을 시도할 것이며 이를 위해 차단된 경로가 아닌 다른 경로로 우회를 시도할 것이다.

3.2 탐지 방안

3.2.1 반복 접속 탐지

내부에 유입된 에이전트 파일이 외부 공격자의 지령을 받거나 추가적인 파일을 다운로드 하기 위해 외부와 반복적인 접속을 시도한다. 기존 반복접속을 찾기 위한 방법들은 일정한 시간이 똑같이 반복되는 경우 이거나(예: 10분 간격으로 매10분마다 접속 여부를 확인), 발생한 접속을 찾아내어 시간을 계산하여 간격이 비슷한 것을 찾아내었다(예: 접속 간격이 9분, 10분, 11분이면 비슷한 접속으로 간주). 그러나 이러한 방식의 탐지는 접속이 규칙적이지 않은 경우 탐지하기 어렵다. 또한 접속의 주기성을 탐지 당하지 않기 위해 고의적으로 특정 주기를 나타내지 않는 접속을 시도할 수 있다. 따라서 Fig.2.에서와 같이 접속이 불규칙적이든, 규칙적이든 상관없이 이러한 접속 시도를 탐지하기 위해 다음 방안을 제안한다.

1) 네트워크 Flow와 http header 정보들을 토대로 시간대별로 그룹핑(grouping)을 한다. 이때 사용자 클릭에 의한 것이 아닌 것을 찾아내기 위해 referer 값이 없는 것만을 대상으로 한다. 탐지하고자 하는 트래픽은 에이전트에 의해 자동으로 발생되는 트래픽이기 때문이다. referer 값은 웹페이지 이

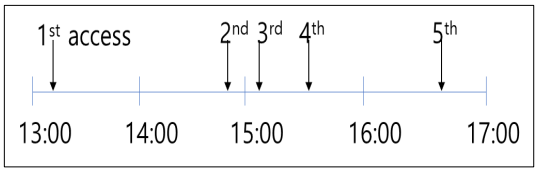

Fig. 2. Irregular Iteration Access

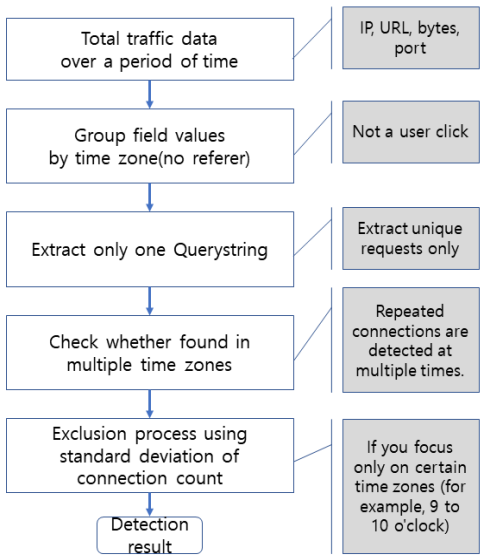

Fig. 3. Flow-chart(Irregular Iteration Access)

동시 이전 페이지의 참조 정보를 나타내므로 이러한 정보가 있는 경우는 사용자가 정상적으로 웹페이지 링크 클릭을 통해 생성된 트래픽이라고 간주하여 생략한다.

2) 그룹핑한 데이터 중 쿼리스트링(QueryString)이 1개의 값만 가지는 것을 추출한다. 쿼리스트링이 그룹핑에서 1개만 존재한다는 것은 외부에 접속을 요청하는 정보 중 유니크한 접속 시도인 것이다. 쿼리스트링이 여러개면 한가지 접속이 반복으로 한 것이 아니라 여러 개의 접속이 발생한 것이고, 여러개의 동일한 접속이 발생한다는 것은 정상적인 사용자에 의한 패턴으로 볼 수 있으므로 제외시킨다. 따라서 다수가 접속하는 패턴이 아니라 공격자에게 지령을 요청하는 트래픽을 찾기 위해 1개의 값만 가지는 것을 추출한다.

3) 위의 수행결과가 몇 개의 시간대에서 발견되는지를 확인하여 탐지 기준을 설정한다. 반복적인 접속을 시도하는 이벤트라면 여러개의 시간대에서 발견될 것이다.(예: 1~2시 발견, 2~3시 발견 등)

4) 위의 3)에서 여러 시간대에 탐지되더라도 업무 시작시간과 같은 특정 시간대에 집중되는 경우는 SW 업데이트와 같은 정상적인 행위일 수 있으므로 각각의 시간대별로 발생한 접속횟수를 표준편차로 계산하여 시간대별로 일정한 횟수를 유지하는 것만을 최종 추출한다.

3.2.2 다중 접속 탐지

공격자가 한 가지 접속만 시도하다가 탐지되면 공격이 무산될 수 있고, 또한 다수의 공격지를 통해 여러 가지 지령을 받을 수 있으므로 이를 위해 다중 접속을 시도할 수 있다. 다중으로 접속을 시도하는 트래픽을 탐지하기 위해 다음 방안을 제안한다.

1) 3.2.1 반복 접속 탐지에서와 같이 http header 정보 중 referer가 없는 데이터를 기준으로 그룹핑을 하고 쿼리스트링 정보가 유일한 것을 추출한다. 쿼리스트링이 유일하면 일반적으로 많이 나타나는 접속 유형을 제외시킬 수 있다.

Fig. 4. Multiple Access

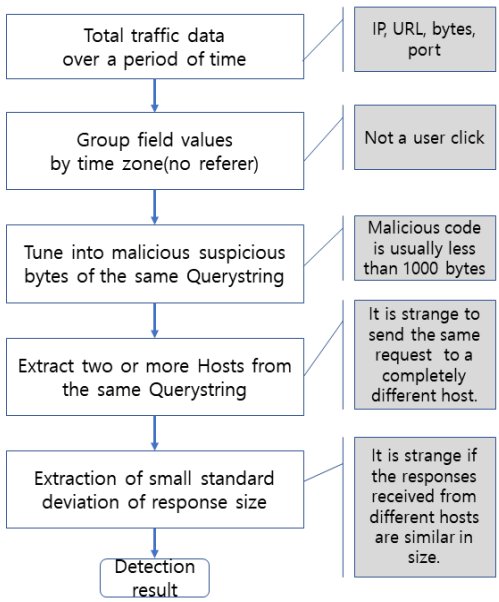

Fig. 5. Flow-chart(Multiple Access)

2) 쿼리스트링 정보가 유일한 것을 추출한 후, 그 중 Fig.4.에서와 같이 Host의 정보가 2개 이상인 것을 추출한다. 쿼리스트링은 외부에 요청 정보를 보내는 정보를 나타내므로 같은 요청 정보를 갖는데 2개 이상의 목적지로 접속하려는 것을 찾는다.

3) 위의 2)에서 추출된 2개 이상의 host를 갖는 접속 내역 중 응답 byte 크기의 표준편차가 작은 것을 추출한다. 쿼리스트링 요청정보가 동일하면서 서로 다른 Host로 접속을 요청했는데, 돌아오는 응답 크기 값이 표준편차가 작아 유사하다면 동일한 요청 정보를 회신 받는 것이라고 추정할 수 있다. 즉, 전체 트래픽 중 독특한 요청 정보를 가지는데, 서로 다른 Host로 동일하게 요청을 하며, 이에 대한 응답 크기 값마저 비슷한 것을 찾는다.

3.2.3 우회시도 접속 시도 탐지

내부에 유입된 에이전트는 외부 통신이 탐지되면 공격에 실패하게 되므로 탐지되지 않기 위해 여러 가지 노력을 한다. 이중 IP정보와 Host정보를 다르게 변조하여 겉으로 보이는 http header 정보는 잘 알려진 사이트 또는 안전한 사이트로 접속을 하는 것처럼 보이면서, 실제로는 공격자 IP로 접속을 시도할 수 있다. 이러한 트래픽을 탐지하기 위해 다음 방안을 제안한다.

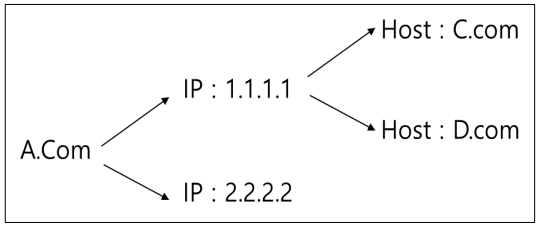

Fig. 6. Bypass Attempt Access

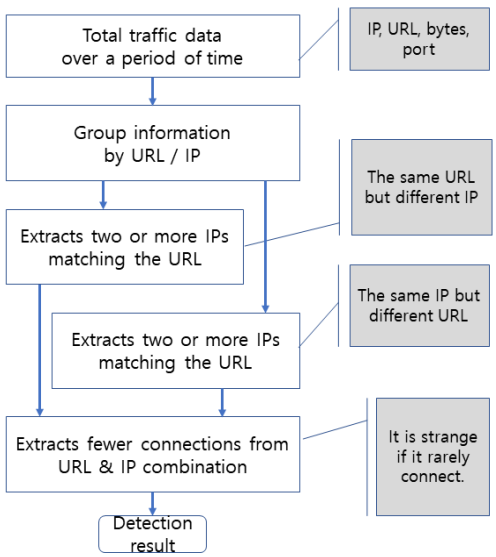

Fig. 7. Flow-chart(Bypass Attempt Access)

1) Flow와 http header 정보들 중 IP와 URL(Host) 정보를 기준으로 그룹핑을 한다. 그룹핑을 하면 IP별로 매칭된 Host 정보를 알 수 있다.

2) URL(Host)에 매칭된 IP가 2개 이상인 것을 추출한다. URL은 각각 고유한 서비스 IP를 가지고 있다. 그런데 같은 URL(Host) 정보를 가지고 있으면서 IP가 다르다면 의심스럽다. 즉, A.com의 IP주소로 1.1.1.1이 있는데, 같은 A.com 이면서 2.2.2.2 인 것을 탐지한다.

3) 그런데 2)의 경우는 구글, 네이버 등과 같은 대형 사이트의 경우 다수 사용자에게 서비스를 하기 위해 여러개의 웹서버와 서비스 환경을 운영할 수 있으며 이럴 경우 같은 Host 정보에 여러개의 IP가 있을 수 있다. 따라서 2)번 결과만으로는 부족하므로 추가적으로 탐지 방안을 적용한다. 이번에는 IP를 기준으로 URL(Host)가 2개 이상인 것을 추출한다. 즉, 목적지 IP가 1.1.1.1 인데 하나는 c.com이고 다른 하나는 같은 1.1.1.1 주소인데 d.com인 경우를 찾는다.

4) 3)의 경우도 소규모 사이트들이 웹호스팅 업체 등을 이용하여 서비스 하는 경우 각각 다른 URL임에도 불구하고 같은 목적지 IP를 가질 수 있다. 따라서 이를 보완하기 위해 Fig.7.과 같이 2)번의 결과와 3)번의 결과를 조합하고 추가로 접속 횟수가 적은 것을 찾는다. 접속자 수가 많으면 다수의 사용자가 접속하는 사이트이므로 정상일 가능성이 높다. 따라서 접속자 수가 적은 것을 추출한다. 즉, 같은 Host인데 목적지 IP 가 다른 경우를 찾고, 여기서 추출된 IP를 토대로 이번에는 같은 IP인데 Host정보가 2개 이상인 것을 찾은 다음, 마지막으로 접속 횟수(접속자수)가 일정량 이하인 것을 찾는다.

3.2.4 우회시도 접속 가능 탐지

3.2.3에서는 탐지되지 않기 위해 헤더 정보를 변조하여 우회를 시도하는 접속을 탐지하는 방안을 제안하였다. 3.2.4에서는 탐지가 되어 차단 조치가 된 경우 아직 내부적으로 해당 트래픽을 발생시키는 에이전트가 완전히 제거되지 못해 다른 경로로 접속을 다시 시도할 가능성이 있는 트래픽을 탐지한다. 에이전트는 외부와의 접속을 지속적으로 시도하는데 해당 경로가 차단되어 접속이 일정기간 이뤄지지 않으면 접속 경로를 바꿔 다른 외부의 공격자 주소로 접속을 시도할 수 있다. 이러한 경우를 탐지하기 위해 다음 방안을 제안한다.

1) 방화벽 접속 트래픽을 모두 모아 IP, Port 등의 정보로 그룹핑을 실시한다. 그룹핑을 통해 접속 내역 현황 파악이 가능하다.

2) 그룹핑된 내역 중, 방화벽 이벤트의 차단(deny) 로그를 추출하고, 해당 접속 내역의 접속 시도 건수가 특정 횟수(예: 50회 이상/1일) 이상인 것을 별도로 추출한다.

3) 차단된 목적지로 하루에 특정 횟수 이상 접속을 시도한다는 것은 정상적인 사용자에 의한 접속이나 실수가 아니라, 반드시 에이전트에 의한 접속 시도이며, 이러한 트래픽은 외부로 연결되는 경로는 차단되었지만 실제로 트래픽을 발생시키는 에이전트 자체가 조치되지 않았음을 의미한다. 따라서 해당 에이전트가 경로를 차단되지 않은 주소로 재접속을 하여 추가 공격이 이뤄지지 않도록 사전에 트래픽을 발생시키는 에이전트를 조사하여 제거하는 노력이 필요하다.

3.3 탐지 한계

본 논문에서는 사전에 유입된 에이전트를 찾기 위해 몇가지 전제사항을 마련하였다. 외부와 반드시 반복적인 통신을 시도할 것이며, 다중 접속과 우회시도를 하는 것을 전제로 하여 해당 방법을 탐지하기 위한 방안을 제안하였다. 그러나 반복을 계산하기 위한 범위를 조절하는 것에 따라 탐지 내역은 달라질 수 밖에 없다.

본 논문에서는 하루의 데이터를 토대로 분석하여 시간대별로 나타나는 접속 이벤트를 분석하였으나, 공격자의 에이전트가 하루에 1회 이하로 접속을 시도하게 제작되었거나 접속 시도마다 요청 정보를 변경할 경우 탐지에서 제외 될 수 있다. 또한 대량의 트래픽이 발생하는 대규모 조직에서는 위의 탐지 방법에 의해 추출된 트래픽이 매우 많을 수 있다. 이러한 경우는 화이트리스트 관리를 통해 추출되는 모수를 조절해가는 노력이 필요 하다. 따라서 본 논문에서는 변종 악성코드가 탐지 장비를 우회하여 침입하였고 하루 24시간 중 3개 이상의 다수 시간대에서 탐지 회피수단(불규칙적 접속, 다중우회 접속)을 시도하며 외부 공격자와 통신하려는 트래픽을 탐지하는 것에 집중하여 제안되어 있다.

IV. 실험 결과 및 분석

4.1 실험 데이터

본 논문의 탐지 방안을 실험하기 위해 오픈소스 기반으로 데이터 수집과 분석을 할 수 있는 기반환경을 구축하였다. 수집부분은 다양한 종류의 데이터 실시간 수집이 가능한 FLUME, 데이터의 저장은 대용량 파일 시스템 하둡의 HDFS, 데이터 처리 부분은 메모리 기반 데이터 처리 SPARK를 사용하였으며, 분석부분은 HDFS의 파일시스템 기반 데이터를 SQL형태로 분석할 수 있는 HIVE를 사용하였다.

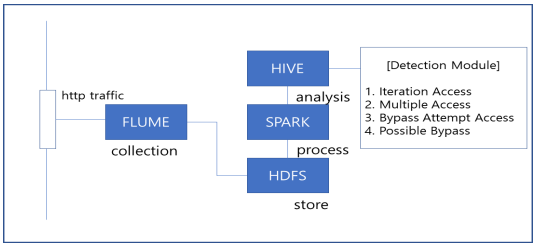

Fig. 8. Experiment Environment

실험에 사용한 데이터는 국내 S기업의 2016년 11월 15일부터 12월 30일 까지 약 6주 동안 발생한 실제 인터넷 접속 트래픽 중, Flow와 Http 통신 일부를 사용하여 실험하였다. 데이터는 1일 평균 약740GB가 발생하였으며, 실험기간 전체 동안 누적된 데이터양은 약31TB 규모이다.

실험에는 Flow와 http header 정보 중 전문가 검토를 통해 15개 항목을 선정하여 활용하였다.

Table 1. Selected Header Field

No.	Field	Type
1	Start Time	Date
2	Application Name	String
3	Source IP	String
4	Source Port	Integer
5	Source Bytes	Integer
6	Destination IP	String
7	Destination Port	Integer
8	Destination Location	String
9	Destination Bytes	Integer
10	Http Method	String
11	Http Host	String
12	Http Response Code	String
13	Http User Agent	String
14	Http Referer	String
15	Http QueryString	String

4.2 실험 결과 및 분석

실험 기간 동안 매일 약 740GB의 데이터를 분석하였으며, 6주 동안 총 31TB를 분석한 결과 에이전트에 의한 접속으로 총203건이 탐지 되었다. 203건은 기존에 보유한 보안장비에서 중복 탐지된 내역을 제외한 신규 탐지 건수이며, 이중 약 29%에 해당하는 59건이 악성으로 확인되거나 비업무성 활동으로 확인되었다.

각 탐지 내역을 분석한 결과 C&C, 랜섬웨어IP, 악성코드 유포지와 통신한 악성 사이트 접속 4건은 즉시 접속자에게 통보 및 차단 조치되었다. 4건의 탐지 건 중 2건은 비표준 port를 이용하여 C&C통신을 하며 다량의 스팸메일을 발송하는 악성코드에 의한 것으로 확인되었으며 '반복접속탐지' 모듈에 탐

Table 2. Result of Verification

division	content	Number of detections
Malicious site access	C&C	2
	Distribute malicious code	1
	Ransomware IP	1
Non-business site access	game	6
	chatting	1
	stock	9
	adware	1
Risk access tool (Internal use restriction tool)	P2P(Torrent)	9
	Remote Access View	4
Access blocked sites	Attempted to access blocked site more than 50 times a day	21
Other (not confirmed)	404 Not found	4

지되었다. 또 다른 1건은 Conficker.Worm 감염에 의한 것으로 1:N의 통신이 지속적, 반복적으로 발생하며 추가적인 악성코드를 다운로드하기 위한 시도를 수행하다가 '반복접속탐지'와 '다중접속탐지'에 걸렸다. 마지막 1건은 트래픽 추적 결과 원인이 되는 악성코드를 발견하는데 실패하였으나 PC 사용자가 인지하지 못한 채 외부 프록시(proxy) 서버와 랜섬웨어 유포에 연관된 IP 접속을 시도하다가 '반복접속탐지' 모듈에 탐지되었다. 위의 트래픽을 발생시킨 에이전트(악성코드)들이 위와 같은 방식으로 탐지 당하지 않기 위해 주기성을 줄이고 사이트(host/ip)를 바꿔가며 접속을 시도하는 형태로 변형되더라도 '우회접속탐지' 방법에 의해 탐지가 가능할 것으로 예상한다.

게임, 주식과 같은 비업무성 사이트에 지속 접속한 에이전트들과 사내 사용이 제한된 위험성 도구들은 현재 악성코드로 분류되지 않았으나 해킹에 활용될 잠재적 위험 요소로 분류되어 담당자들에게 통보되었다.

차단된 사이트 접속 건은 이미 방화벽에서 해당 사이트가 악성으로 판단되어 차단된 사이트임에도 불구하고 지속적으로 접속을 시도하여 원인이 해결되지 않는 접속 건으로, 해당 PC 사용자에게 추가적인 확인 조치가 통보되었다.

그밖에 http 응답코드가 '404 Not found'에 해당하는 존재하지 않는 페이지에 수백차례 접속을 시도하는 비정상적인 트래픽이 탐지되어 조치하였다.

203건의 탐지 내역 중 59건을 제외한 나머지 144건은 내부 시스템간의 정상통신이거나, 모바일 기기의 앱을 위한 통신 등으로 확인되었다. 이러한 부분은 지속적으로 화이트리스트 처리하여 관리하거나 별도의 항목으로 분류하여 관리하는 추가적인 작업이 요구된다.

V. 결 론

기존 탐지 연구들이 실시간 발생하는 이벤트에 대해 탐지룰을 걸어 탐지하였다면, 본 논문에서는 하루 동안 발생한 트래픽 전체를 대상으로하여 규칙성을 띄지 않더라도 반복적으로 접속하는 내역을 찾아내는 방법 등 새로운 방안을 제시하였다. 실제 발생한 데이터를 바탕으로 실험한 결과에서도 기존에 보유한 여러 탐지 장비에서 탐지되지 못한 악성 트래픽 등 의미 있는 결과를 찾아냈다. 하지만 데이터가 많아질경우 내부 시스템간 통신과 같이 정상적인 트래픽으로 확인되는 내역이 많아져 분석 건수가 증가하게 되므로 이를 화이트리스트 처리해야하는 작업이 필요했다.

향후에는 화이트리스트를 안전하고 쉽게 관리하는 방안과, 최근 사용량이 증가하고 있는 https 트래픽에 대한 탐지 방법에 대해 추가적인 연구를 진행하고자 한다. 또한 현재 실험 환경인 하둡(hadoop) 저장소인 HDFS(Hadoop File System)의 느린 속도 문제를 개선하기 위해 NoSQL, 검색엔진 기반 데이터 플랫폼 기술 등을 활용하여 실제 보안관제에 적용하여 기술기반으로 보안관제 역량을 높일 수 있는 방안을 제시할 것이다.

References

[1] IDC, 3rd platform Digital Transformation, https://www.idc.com/promo/thirdplatform/digitaltransformation

[2] IDG, Top 5 cybersecurity facts, figures, statistics for 2018, https://www.csoonline.com/article/3153707/security/top-5-cybersecurity-facts-figures-and-sta

tistics.html
〔3〕 Boannews, http://www.boannews.com/media/view.asp?idx=69212
〔4〕 Lockheedmartin, Cyber Kill Chain, https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html
〔5〕 AVTEST(The Independent IT-Security Institute), Malware Statistics, https://www.av-test.org/en/statistics/malware/
〔6〕 Kirti Mathur and Saroj Hiranwal, "A Survey on Techniques in Detection and Analyzing Malware Executables", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 4, 2013.
〔7〕 PAYLOAD SECURITY, "Hybrid Analysis - Innovative Technology", https://www.payload-security.com/technology/hybrid-analysis
〔8〕 Jan Goebel, Thorsten Holz, "Rishi: Identify Bot Contaminated Hosts by IRC Nickname Evaluation", USENIX Hot Bots. pp.4-9, July. 2007.
〔9〕 Marina Thottan and Chuanyi Ji, "Anomaly Detection in IP Networks", IEEE Transaction On Signal Processing, Vol. 51, No. 8, Aug. 2003.
〔10〕 K. Illgun, R. Kemmerer, Phillip A. Porras, "State Transition Analysis : A rule-based intrusion detection approach," IEEE Transaction On Software Engineering, pp.181-199, Mar. 1995.
〔11〕 Shen Maying, Jiang Xinghao, Sun Tanfeng, "Anomaly detection based on Nearest Neighbor search with Locality-Sensitive B-tree," Neurocomputing, Vol. 289, pp.55-67, May. 2018.
〔12〕 Tonejc Jernej, Kobekova Alexandra, "Machine Learning Methods for Anomaly Detection in BACnet Networks," Journal Of Universal Computer Science, Vol. 22, No 9, pp.1203-1224, 2016.
〔13〕 Liu, Weixin, Zheng, Kangfengm "Flow-based Anomaly Detection Using Access Behavior Profiling and Time-sequenced Relation Mining," KSII Transactions On Internet And Information Systems, Vol. 10, Issue 6, pp.2781-2800, June. 2016.
〔14〕 Siwoon Son, Myeong-Seon Gil, "Anomaly Detection of Hadoop Log Data Using Moving Average and 3-Sigma," KIPS Tr. Software and Data Eng, Vol. 5, No. 6, pp.283-288, 2016.

〈저자소개〉

조 영 민 (Young Min Cho) 정회원
2002년 2월: 고려대학교 물리학과 졸업
2017년~현재: 고려대학교 정보보호대학원 석사 과정
〈관심분야〉 보안인식, 머신러닝, 인공지능, 빅데이터, 위협정보

권 헌 영 (Hun Yeong Kwon) 종신회원
1992년 2월: 연세대학교 법학과 졸업
1998년 2월: 연세대학교 법학과 석사
2005년 2월: 연세대학교 법학과 박사
2015년 9월~현재: 고려대학교 정보보호대학원 부교수
〈관심분야〉 정보보호법 및 정책, 정보통신법 및 정책, 사이버법률, 인터넷규제, 전자정부